Introduction to Probability and Statistics

Sec 5: Statistical Inference

A Guessing Game

- This chapter presents the key ideas of Statistical Inference.
- Probably the area of statistics other scientists are most familiar with.

  (Take a surgeon, a biologist, a meteorologist: they'll all be familiar with t-tests, p-values, and similar concepts we'll discuss in this chapter)

A Guessing Game

- This chapter presents the key ideas of Statistical Inference.
- Probably the area of statistics other scientists are most familiar with.

  (Take a surgeon, a biologist, a meteorologist: they'll all be familiar with t-tests, p-values, and similar concepts we'll discuss in this chapter)
- However, before seeing what Statistical Inference is about, let's play a "Guessing Game"!

Python Notebook

Chapter Overview

- What you've just done is carrying out some Statistical Inference: using a finite sample of data to *infer* (deduce) properties of an unknown parameter of a population.

- Specifically, you have used some data to **estimate** the mean μ of a Normal distribution, and used your guess to:
  - Reject/accept some **hypothesis** about μ.
  - Provide **an interval of plausible values** for μ.

Chapter Overview

- What you've just done is carrying out some Statistical Inference: using a finite sample of data to *infer* (deduce) properties of an unknown parameter of a population.

- Specifically, you have used some data to **estimate** the mean μ of a Normal distribution, and used your guess to:
  - Reject/accept some **hypothesis** about μ.
  - Provide **an interval of plausible values** for μ.

- Indeed, in this chapter we will study:
  1. Estimators
  2. Confidence Intervals
  3. Hypothesis Testing

  A tool will be key throughout: The Central Limit Theorem.

Statistical Inference

In statistics we are often interested in estimating a summative property of a population – *e.g.*, the mean or the variance:

$\longrightarrow$ We use a **sample** to conclude about the **population**.

Example: Interested in the mean height of Cameroonian ladies between 30 and 40. *On one day*, we collect this info for *some* ladies on the street.

- **Population**: All Cameronian ladies between 30 and 40.
- **Sample**: The Cameroonian ladies we collect data from.

Parameters and Statistics

Parameters

Fixed, unknown values of population.

e.g. Mean (μ), variance (σ^2)

Sample Statistics

A value calculated from the sample.

e.g. Sample mean ($\bar{x}$), sample variance (s^2)

Sample Statistics and Sampling Distribution

Let $x = (x_1, \ldots, x_n)$ be a random sample collected from the population.

- A **Sample Statistic** $T(x)$ provides a summary of the values in x. Sample statistics are random variables: Their value will vary from one sample to the other (think of the Cameroonian height example).

- The probability distribution of $T(x)$ is called **Sampling Distribution**. It describes how the statistic varies between samples.

For example, the statistic

$$T(x) = \frac{x_1 + \ldots + x_n}{n} = \bar{x}$$

can be used to *draw inference* about the true but unknown mean μ.

How Does Sample Size affect Sampling Distribution?

Consider the Cameroonian ladies' height example. We have:

- Variable of interest: $X =$ height.
- Parameter of interest: $\mu = \mathbb{E}[X]$, the true average height.
- Sample Statistic: $T(x) = \bar{x}$, the sample mean.

Imagine to repeat the data collection many times, say 1000 times, for each of the following values of n. How do you think the sampling distribution changes?

- $n = 5$, so $T(x) = \frac{1}{5}\sum_{i=1}^{5} x_i$
- $n = 50$, so $T(x) = \frac{1}{50}\sum_{i=1}^{50} x_i$
- $n = 500$, so $T(x) = \frac{1}{500}\sum_{i=1}^{500} x_i$

??

Central Limit Theorem

The Central Limit Theorem (CLT) answers the above question. It tell us that:

Whatever distribution we sample the x_i from, as long as n is sufficiently large, the distribution of the sample mean $T(x) = \bar{x}$ will be approximately normal:

$$\bar{X} \;\dot{\sim}\; N\left(\mu, \frac{\sigma^2}{n}\right).$$

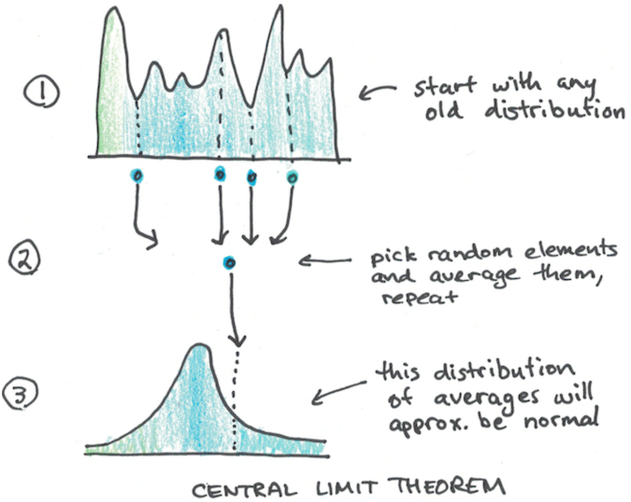
We can now answer the previous question. If we takes 1000 samples of size n and average each of them, then the 1000 sample means:

1. Will fluctuate around the original mean μ $\quad(\mathbb{E}[\bar{X}] = \mu)$.
2. If we increase n, they'll be more concetrated towards μ $\quad(\mathbb{V}\text{ar}[\bar{X}] = \sigma^2/n)$
3. Will approximately follow a normal distribution $\quad(\bar{X} \;\dot{\sim}\; N(\ldots))$

Central Limit Theorem: A few Remarks

- The statement is really powerful: whatever the original distribution, the averaging process will always return an (approximately) normal distribution.
- The approximation gets better and better the larger n is.
- As a rule of thumb, $n \geq 30$ is often considered large enough for the approximation to hold (although it really depends on how far-from-normal the original distribution is!)
- If the original distribution is already normal, then also $\bar{X}$ will be exactly normal.

Central Limit Theorem: Visualisation



Central Limit Theorem: Formal Statement

The proper statement of the CLT is given in terms of n independent RVs, all identically distributed, rather than n independent samples of one RV.

The acronym $i.i.d.$ is used to abbreviate *independent and identically distributed*.

Central Limit Theorem (CLT)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Let:

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n}.$$

Then, for n sufficiently large, the distribution of $\bar{X}$ is *approximately* normal:

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \;\dot{\sim}\; N(0,1).$$

Statistical Estimators

Estimators

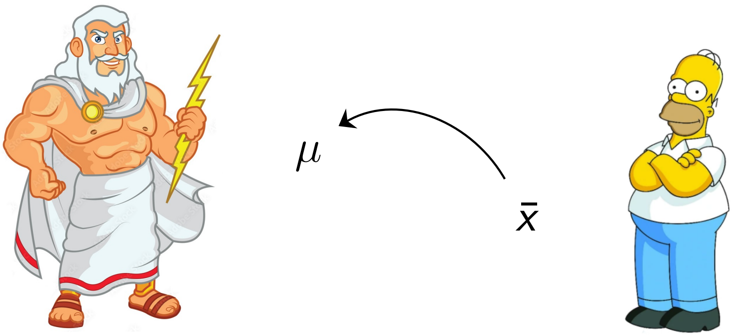
When a sample statistic is used to estimate a parameter θ of a population, the statistic is referred to as an **estimator**.

For example, the sample mean $\bar{x}$ is an estimator of the population mean μ: We do **not** know μ. We use $\bar{x}$ to estimate it.

Estimators

When a sample statistic is used to estimate a parameter θ of a population, the statistic is referred to as an **estimator**.

For example, the sample mean $\bar{x}$ is an estimator of the population mean μ: We do **not** know μ. We use $\bar{x}$ to estimate it.



Desirable Properties

- An estimator $T(x)$ of a parameter θ is **unbiased** if the expected value of the sampling distribution of T is equal to θ.

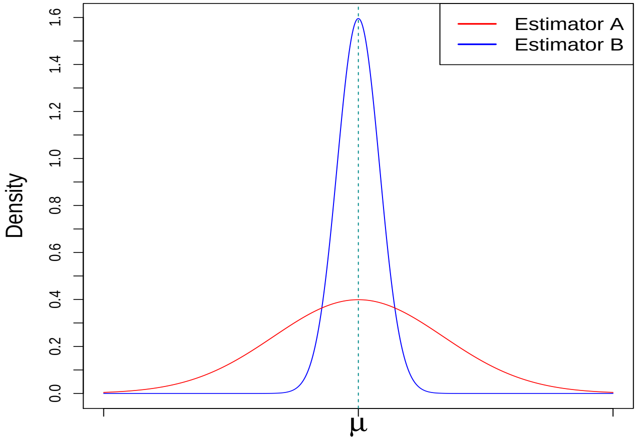
$$\mathbb{E}[\,T(x_1, \ldots, x_n)] = \theta \quad \Longrightarrow \quad \bar{X} \text{ unbiased estimator of } \theta$$

- The **standard error** of an estimator $T(x)$ is the standard deviation of the sampling distribution of T.

$$SE_{T(x)} = \mathrm{Std}[T(x)]$$

- Generally speaking, it is desirable for an estimator to be unbiased and to have a small standard error.

Comparing Estimators



Both estimators are unbiased. However, B is preferable to A due to its smaller SE.

Examples of Estimators

Given a sample from n iid rvs $X_1, \ldots, X_n$ with mean μ and variance σ^2:

- The sample mean $\bar{X}$ is an unbiased estimator of the population mean μ (since $\mathbb{E}[\bar{X}] = \mu$)
- The standard error of $\bar{X}$ is $SE_{\bar{X}} = \sigma/\sqrt{n}$ (since $\mathbb{V}\text{ar}[\bar{X}] = \sigma^2/n$)

Examples of Estimators

Given a sample from n iid rvs $X_1, \dots, X_n$ with mean μ and variance σ^2:

- The sample mean $\bar{X}$ is an unbiased estimator of the population mean μ (since $\mathbb{E}[\bar{X}] = \mu$)
- The standard error of $\bar{X}$ is $SE_{\bar{X}} = \sigma/\sqrt{n}$ (since $\mathbb{V}\mathrm{ar}[\bar{X}] = \sigma^2/n$)

Moreover, it can be shown that the sample variance s^2 is an unbiased estimator of the population variance σ^2:

$$\mathbb{E}\left[s^2\right] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2$$

Examples of Estimators

This explains *why* we defined the sample variance with a $\frac{1}{n-1}$ instead of $\frac{1}{n}$ in front of the sum.

Indeed, the last formula can be rewritten as:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n-1}{n}\,\sigma^2.$$

Hence, $\tilde{s}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$ is a *biased* estimator of σ^2.

The sample variance s^2 is instead an unbiased estimator of σ^2.

Confidence Intervals

Confidence Intervals

Think of the game we played. I give you n independent samples from a distribution: can you provide plausible guesses of the original mean?

$$x_1, \ldots, x_n \text{ observed: } \quad \text{what could } \mu \text{ be?}$$

- A **confidence interval (CI)** for a population parameter θ is an interval of values that we are "confident" to contain the true parameter θ.

Confidence Intervals

Think of the game we played. I give you n independent samples from a distribution: can you provide plausible guesses of the original mean?

$$x_1, \dots, x_n \text{ observed: } \quad \text{what could } \mu \text{ be?}$$

- A **confidence interval (CI)** for a population parameter θ is an interval of values that we are "confident" to contain the true parameter θ.
- Historically, the level of confidence is denoted $1 - \alpha$. We want this close to 1, so α is usually small.

$$\begin{aligned} \alpha &= 0.10 \quad \rightarrow \quad 90\% \text{ confidence} \\ \alpha &= 0.05 \quad \rightarrow \quad 95\% \text{ confidence} \qquad \text{(very common)} \\ \alpha &= 0.01 \quad \rightarrow \quad 99\% \text{ confidence} \end{aligned}$$

Confidence Interval for the Mean: Large Sample

- Here we'll focus on CIs for the mean (by far the most common case). So let's see how, starting from a sample $x_1, \ldots, x_n$, we can compute a $100(1-\alpha)\%$ CI for the population mean.

Confidence Interval for the Mean: Large Sample

- Here we'll focus on CIs for the mean (by far the most common case). So let's see how, starting from a sample $x_1, \ldots, x_n$, we can compute a $100(1-\alpha)\%$ CI for the population mean.

- Let's understand the reasoning in the 95% CI case ($\alpha = 0.05$), and generalise after.

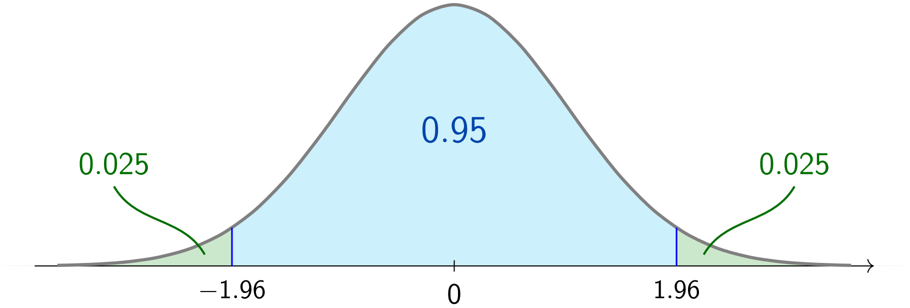
- We'll assume a large sample, so that the CLT applies:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Confidence Interval for the Mean: Large Sample

We need to estimate μ with $\bar{x}$. Recall that $\bar{x}$ varies sample to sample.

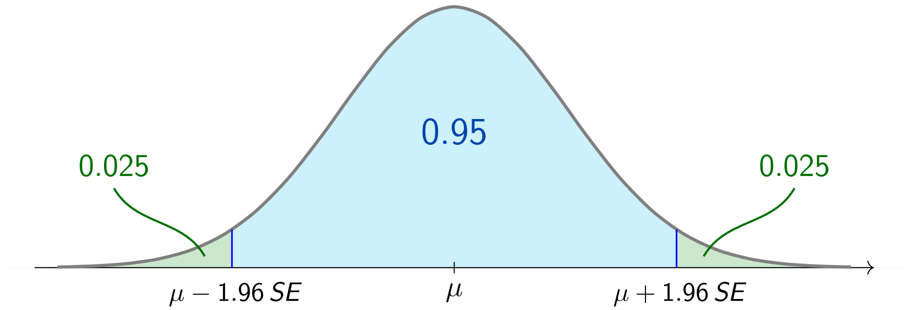
- The different $\bar{x}$ are distributed normally around μ, with std $\sigma/\sqrt{n}$.
- So, 95% of sample means $\bar{x}$ will be between: $\mu \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$



Confidence Interval for the Mean: Large Sample

We need to estimate μ with $\bar{x}$. Recall that $\bar{x}$ varies sample to sample.

- The different $\bar{x}$ are distributed normally around μ, with std $\sigma/\sqrt{n}$.
- So, 95% of sample means $\bar{x}$ will be between: $\mu \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$



Confidence Interval for the Mean: Large Sample

"Reverting" the reasoning. For any one sample x, we can consider the interval:

$$I(x) = \left[\bar{x} - 1.96 \frac{s}{\sqrt{n}} , \; \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right].$$

Then, we expect 95% of such (random) intervals to contain μ.

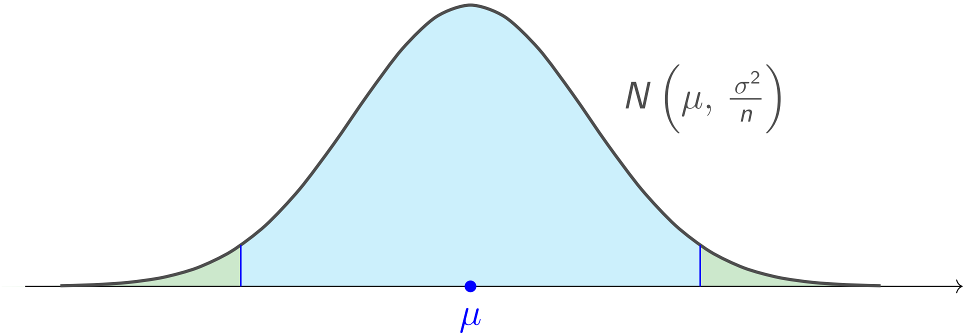
Definition: $I(x)$ is the 95% **Confidence Interval** for the mean, associated with the sample x.

Note: We use the sample mean $\bar{x}$ and sample std s of the sample to compute $I(x)$. We know these, while we don't know the original μ and σ!

Confidence Intervals: Dynamic Visualisation

$$95\% \text{ CI for } \mu: \quad \left[\bar{x} - 1.96\,\frac{s}{\sqrt{n}}\,,\ \ \bar{x} + 1.96\,\frac{s}{\sqrt{n}}\right]$$

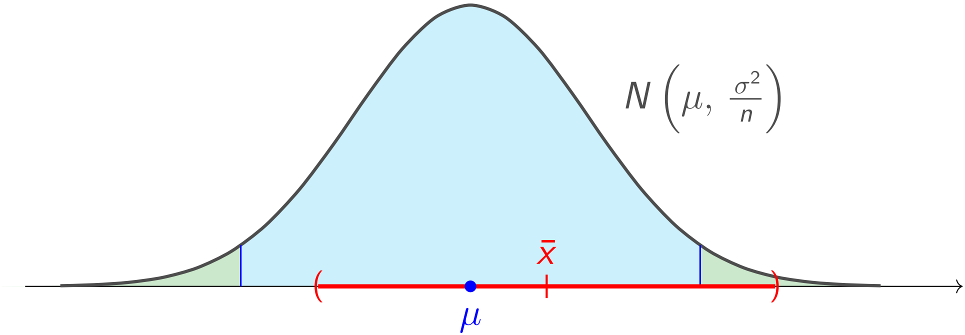
Approximately 95% of them will contain μ.



Confidence Intervals: Dynamic Visualisation

95% CI for μ: $$\left[\bar{x} - 1.96\,\frac{s}{\sqrt{n}}\,,\;\; \bar{x} + 1.96\,\frac{s}{\sqrt{n}}\right]$$

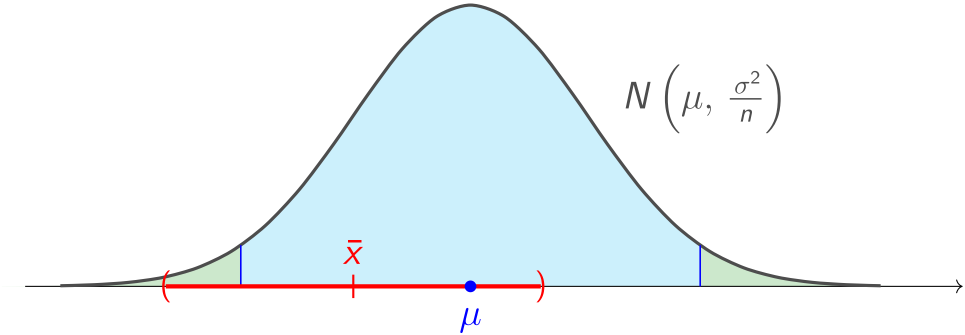
Approximately 95% of them will contain μ.



Confidence Intervals: Dynamic Visualisation

$$95\% \text{ CI for } \mu: \quad \left[\bar{x} - 1.96\,\frac{s}{\sqrt{n}}\,,\;\; \bar{x} + 1.96\,\frac{s}{\sqrt{n}}\right]$$

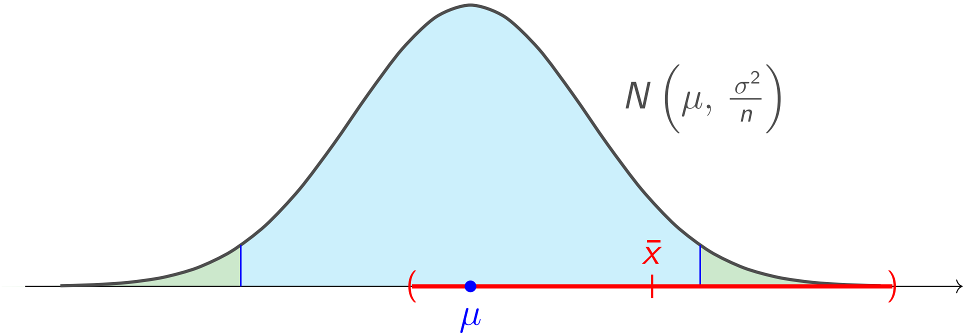
Approximately 95% of them will contain μ.



Confidence Intervals: Dynamic Visualisation

$$95\% \text{ CI for } \mu: \quad \left[\bar{x} - 1.96\,\frac{s}{\sqrt{n}}\,,\;\; \bar{x} + 1.96\,\frac{s}{\sqrt{n}}\right]$$

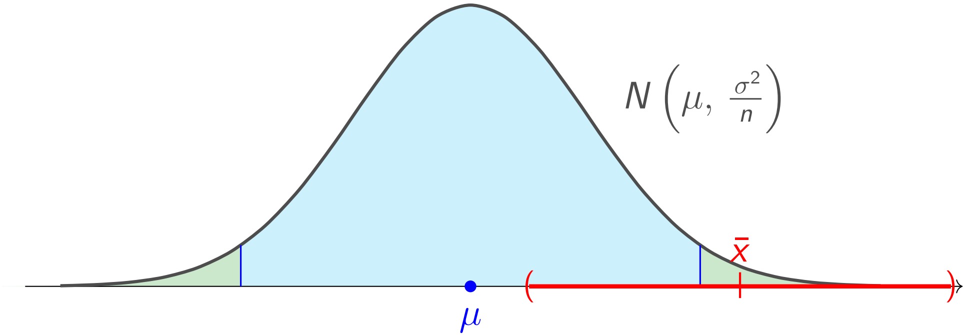
Approximately 95% of them will contain μ.



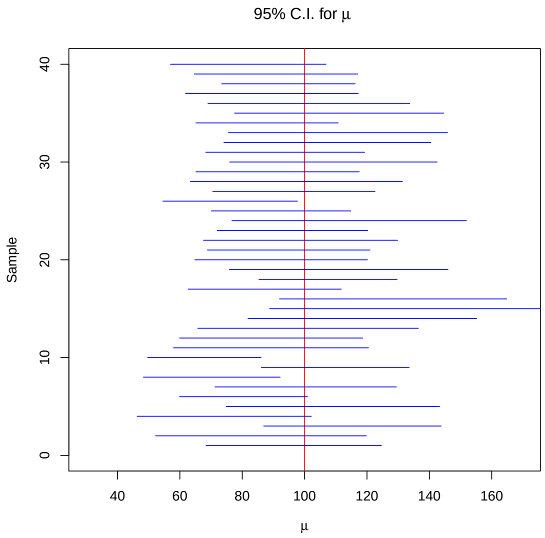
Confidence Intervals: Dynamic Visualisation

$$95\% \text{ CI for } \mu: \quad \left[\bar{x} - 1.96 \frac{s}{\sqrt{n}} , \;\; \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right]$$

Approximately 95% of them will contain μ.



Confidence Intervals: Static Visualisation



A 95% CI must be such that:

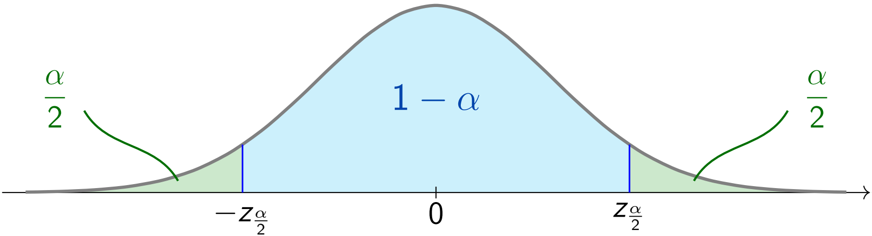
If several samples are drawn
(40 in the picture)

95% of the associated CIs
contain the true parameter.
($\mu = 100$ in the picture)

Confidence Intervals: General Case

- If we want a confidence level $100(1-\alpha)\%$, then we just need to replace 1.96 with the relevant quantile of a standard normal.
- Call z_γ the standard normal value having exactly γ area to its right: $\mathbb{P}(Z > z_\gamma) = \gamma$.
- Then, the $100(1-\alpha)\%$ CI for μ is

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$



Typical Values

Relevant standard normal quantiles

$$\alpha = 0.10 \quad \rightarrow \quad z_{0.05} = 1.645$$
$$\alpha = 0.05 \quad \rightarrow \quad z_{0.025} = 1.96$$
$$\alpha = 0.01 \quad \rightarrow \quad z_{0.005} = 2.576$$

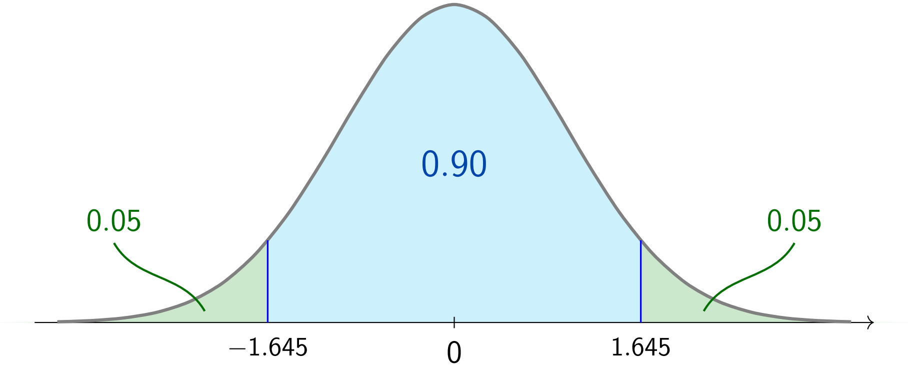
Confidence Intervals

$$90\% \text{ CI for } \mu: \quad \bar{x} \pm 1.645 \text{ SE}$$
$$95\% \text{ CI for } \mu: \quad \bar{x} \pm 1.96 \text{ SE}$$
$$99\% \text{ CI for } \mu: \quad \bar{x} \pm 2.576 \text{ SE}$$

where the standard error is always computed as $\text{SE} = \frac{s}{\sqrt{n}}$.

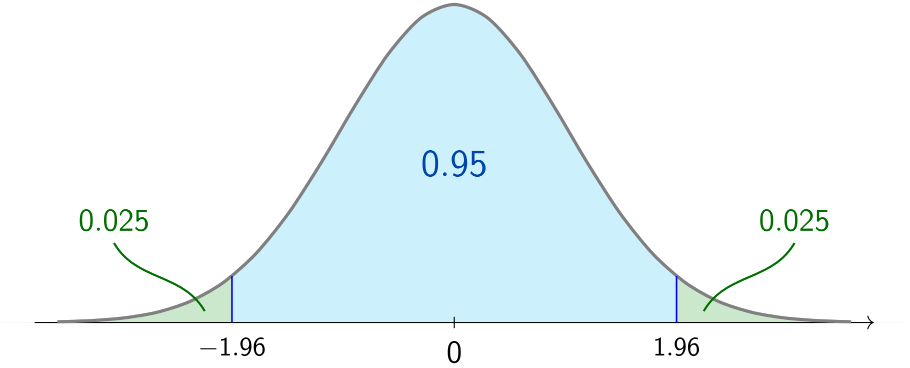
Typical Values

$$90\%\ \text{CI:}\quad \bar{x} \pm 1.645\,\frac{s}{\sqrt{n}}$$



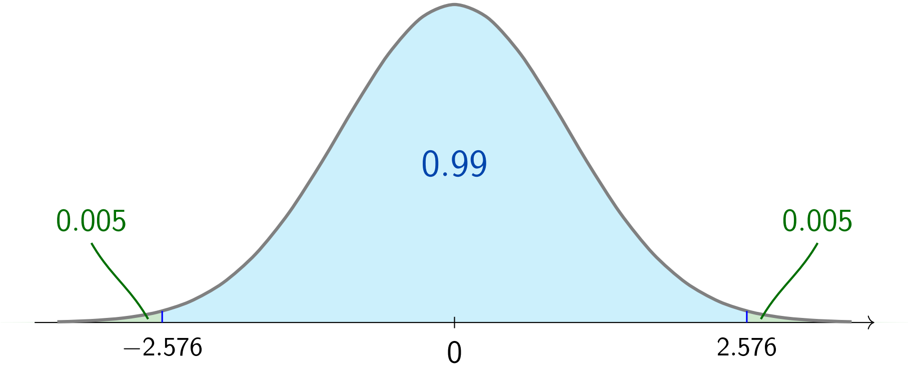
Typical Values

$$95\%\ \text{CI:} \quad \bar{x} \pm 1.96\,\frac{s}{\sqrt{n}}$$



Typical Values

$$99\%\ \text{CI:}\quad \bar{x} \pm 2.576\,\frac{s}{\sqrt{n}}$$



Confidence Intervals: Example

A random sample of 100 observations from a non-normally distributed population has a mean of 83.2 and a standard deviation of 6.4.

- Find a 95% confidence interval for the population mean μ and interpret the interval.
- Find a 99% confidence interval for μ.
- Comment on the width of the intervals.

Confidence Intervals: Example

A random sample of 100 observations from a non-normally distributed population has a mean of 83.2 and a standard deviation of 6.4.

- Find a 95% confidence interval for the population mean μ and interpret the interval.

$$\mathrm{SE} = \frac{s}{\sqrt{n}} = 0.64 \implies \quad 95\%\ \mathrm{CI} = \begin{cases} \bar{x} \pm 1.96\,\mathrm{SE} \\ 83.2 \pm 1.254 \\ [81.946,\ 84.454] \end{cases}$$

- Find a 99% confidence interval for μ.

- Comment on the width of the intervals.

Confidence Intervals: Example

A random sample of 100 observations from a non-normally distributed population has a mean of 83.2 and a standard deviation of 6.4.

- Find a 95% confidence interval for the population mean μ and interpret the interval.
- Find a 99% confidence interval for μ.

  99% CI: $\bar{x} \pm 2.576\,\mathrm{SE} = 83.2 \pm 1.649 = [81.55,\ 84.85]$.
- Comment on the width of the intervals.

Hypothesis Testing

Hypothesis Testing

- One of the most widely used statistical procedures.
- Setting: We formulate a hypothesis on a population, from which we have a sample. Does the data carry enough evidence for us to reject the formulated hypothesis, or not?

Examples:

- Dr. Daniel tells you that bikes will pass on the main road at a rate of 2 per minute. You have the impression you often wait longer than that, so would like to test if Dr. Daniel's statement is correct.
- A pharmaceutical company make a new drug and want to test whether or not it is better than a drug currently on the market.

Basic Steps for Testing Hypotheses

While several different hypothesis tests exist (to test hypotheses on parameters, independence, distributions, etc), the structure of all tests is the same.

1. Determine the null and alternative hypotheses.
2. Collect data and compute a relevant test statistic from them.
3. Evaluate how likely the result would be under the hypothesis.
4. Make a decision.

Let's see in more detail what each of these consists in.

Step 1: Determine hypotheses.

There are *always* two hypotheses you need to formulate.

1. Null hypothesis (H_0): usually says that nothing changed or happened. This is what we hope our data will confute.

2. Alternative hypothesis (H_A): the actual research hypothesis. This is what the researcher suspects to be the case, differently from how things currently are or are believed to be.

The researcher needs to have strong evidence before they can reject the null hypothesis in favour of the alternative. Think of the drug example:

$$H_0\text{: Drug is ineffective} \quad vs \quad H_A\text{: Drug works.}$$

Or of a trial situation:

$$H_0\text{: Defendant is innocent} \quad vs \quad H_A\text{: Defendant is guilty.}$$

Example: Determine hypotheses.

The average height of adults in Cameroon in 2010 was 173 cm. You you wish to test whether it has changed in 2025.

- The null hypothesis is that there is no change:

$$H_0\colon \mu = 173 \text{ cm.}$$

Example: Determine hypotheses.

The average height of adults in Cameroon in 2010 was 173 cm. You you wish to test whether it has changed in 2025.

- The null hypothesis is that there is no change:

$$H_0 \colon \mu = 173 \text{ cm}.$$

- Depending on what exactly you wish to test, there are 3 options for the alternative hypothesis:

  (a) Height of Cameroonian people has increased $\rightarrow$ $H_A \colon \mu > 173$ cm

  (b) Height of Cameroonian people has decreased $\rightarrow$ $H_A \colon \mu < 173$ cm

  (c) Height of Cameroonian people has changed $\rightarrow$ $H_A \colon \mu \neq 173$ cm

Step 2: Collect and Summarise the Data

- The decision in a hypothesis test is based on a single number summary of the observed data. This is called the test statistic.

- The test statistic will guide the decision on whether the data supports H_0.

Step 3: Determine the Likelihood of the Result

In order to decide if the observed result supports H_0, we ask:

> If the null hypothesis is true, how likely are we to observe a test statistic as the one we have observed?

This "likelihood" is quantified by the *p*-value.

- The ***p*-value** is computed as the probability of observing a result as or more "extreme" than the one we have observed, *assuming* H_0 *was true*.
- A low p-value is an indication that H_0 may not hold.

Step 4: Make a Decision.

To make a decision, we need to set a threshold on how low a p-value we are prepared to accept, before we reject the hypothesis H_0 as implausible.

This threshold is called the **significance level** of the test and it's denoted α. We can have two cases.

$p < \alpha$: The p-value is small enough to confidently exclude the possibility that our observed data may have happened under H_0, by chance. We reject the null hypothesis and accept the alternative hypothesis.

$p > \alpha$: The p-value is not small enough to rule out chance under H_0. We cannot reject the null hypothesis. (fail to reject H_0)

Step 4: Courtroom Example (readapted)

Let's readapt the courtroom example to a (lighter) context!

$\mathcal{Q}$: Has a student copied a homework?

$$H_0\text{: Student is innocent} \quad vs \quad H_A\text{: Student is guilty.}$$

$p =$ "Probability of the observed script if student was innocent"

$p < \alpha$: Observed evidence would be unlikely if the student was innocent.
$\rightarrow$ we reject H_0 and declare them guilty.

$p > \alpha$: Observed evidence relatively plausible if student innocent.
$\rightarrow$ we fail to reject H_0 and don't declare them guilty.

Type I and Type II errors

Two possible ways to make a wrong decision.

- **Type I Error:** The null hypothesis is true, but we reject it.
- **Type II Error:** The null hypothesis is NOT true, but we do not reject it.

A Type I error is generally considered more serious than a Type II error.

Type I and Type II errors

Two possible ways to make a wrong decision.

- **Type I Error:** The null hypothesis is true, but we reject it.
- **Type II Error:** The null hypothesis is NOT true, but we do not reject it.

A Type I error is generally considered more serious than a Type II error.

Example: A pharmaceutical company is testing a new drug *(H_0: drug ineffective)*

- Type I error: The drug does not work, but the company concludes it does. The drug is rolled out and patients switch from their old medication to this new ineffective medication.
- Type II error: The drug works, but there is not enough evidence in the data to conclude so. Patients continue on their old medication.

Specific Hypothesis Tests

Tests for the Mean

In this module we study hypothesis tests for the mean (or for comparing two means).

$$x_1, \ldots, x_n \quad \rightsquigarrow \quad \mu\,?$$

We study two types of test:

- z-test: Used when σ (population st.dev.) is known
- t-test: Used when σ is unknown.

Tests for the Mean

In this module we study hypothesis tests for the mean (or for comparing two means).

$$x_1, \ldots, x_n \quad \rightsquigarrow \quad \mu\,?$$

We study two types of test:

- z-test: Used when σ (population st.dev.) is known
- t-test: Used when σ is unknown.

The names come from the distribution followed by the test statistic under H_0:

- Standard normal distribution for the z-test
- t distribution for the t-test (a "modification" of the standard normal)

When n is large, the two tests are almost interchangeable.

$\rightarrow$ in practice, just the z-test is used for large n (it's simpler).

Z-Test

Condition: Population standard deviation σ is known.

We have a sample $x_1, x_2, \ldots, x_n$.

1. Tests that the population mean μ is equal to a specific value:

$$H_0\colon \mu = \mu_0.$$

Z-Test

Condition: Population standard deviation σ is known.

We have a sample $x_1, x_2, \ldots, x_n$.

1. Tests that the population mean μ is equal to a specific value:

$$H_0 \colon \mu = \mu_0.$$

2. The test statistic is a z-score:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \,.$$

Z-Test

Condition: Population standard deviation σ is known.

We have a sample $x_1, x_2, \ldots, x_n$.

1. Tests that the population mean μ is equal to a specific value:

$$H_0: \mu = \mu_0.$$

2. The test statistic is a z-score:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \, .$$

Under H_0, the sampling distribution of z is standard normal:

$$\text{If } H_0 \text{ true } \Rightarrow z \sim N(0, 1).$$

Z-Test: One or Two Tailed

3. Is the observed z too extreme for an $N(0,1)$?
   What counts as "extreme" depends on the alternative hypothesis H_A.

$H_A: \mu \neq \mu_0$, **two-tailed test**

$H_A: \mu < \mu_0$ (lower-tailed) **or** $H_A: \mu > \mu_0$ (upper-tailed)

Z-Test: One or Two Tailed

3. Is the observed z too extreme for an $N(0,1)$?
What counts as "extreme" depends on the alternative hypothesis H_A.

$H_A: \mu \neq \mu_0$, two-tailed test

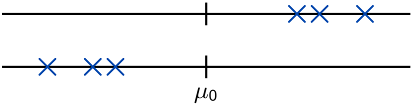
In this case:

- an $\bar{x}$ a lot greter than μ_0 ($z \gg 0$)

OR

- an $\bar{x}$ a lot smaller than μ_0 ($z \ll 0$)

both favour the alternative hypothesis.



$H_A: \mu < \mu_0$ (lower-tailed) **or** $H_A: \mu > \mu_0$ (upper-tailed)

Z-Test: One or Two Tailed

③ Is the observed z too extreme for an $N(0,1)$?
What counts as "extreme" depends on the alternative hypothesis H_A.

$H_A: \mu \neq \mu_0$, two-tailed test

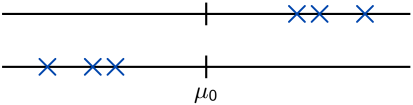
In this case:

- an $\bar{x}$ a lot greter than μ_0 ($z \gg 0$)

OR

- an $\bar{x}$ a lot smaller than μ_0 ($z \ll 0$)

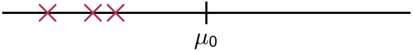
both favour the alternative hypothesis.



$H_A: \mu < \mu_0$ **or** $H_A: \mu > \mu_0$

(lower-tailed) (upper-tailed)

1. H_A: $\mu < \mu_0$. Only $\bar{x}$ lower than μ_0 ($z \ll 0$) are evidence towards H_A.



Z-Test: One or Two Tailed

3. Is the observed z too extreme for an $N(0,1)$?
What counts as "extreme" depends on the alternative hypothesis H_A.

H_A: $\mu \neq \mu_0$, two-tailed test

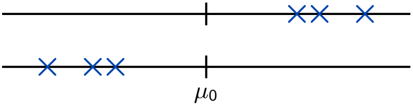
In this case:

- an $\bar{x}$ a lot greter than μ_0 ($z \gg 0$)

OR

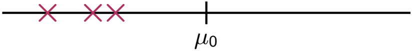
- an $\bar{x}$ a lot smaller than μ_0 ($z \ll 0$)

both favour the alternative hypothesis.

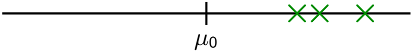


H_A: $\mu < \mu_0$ (lower-tailed) **or** H_A: $\mu > \mu_0$ (upper-tailed)

1. H_A: $\mu < \mu_0$. Only $\bar{x}$ lower than μ_0 ($z \ll 0$) are evidence towards H_A.



2. H_A: $\mu > \mu_0$. Only $\bar{x}$ greater than μ_0 ($z \gg 0$) are evidence towards H_A.



Z-Test: p-value and Final Decision

So, after computing $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$, we calculate the p-value as:

$$
\begin{aligned}
p = \mathbb{P}(Z < -|z| \ \text{ or } \ Z > |z|) \quad & \text{if } \ H_A\colon \mu \neq \mu_0 \\
p = \mathbb{P}(Z < z) \quad & \text{if } \ H_A\colon \mu < \mu_0 \\
p = \mathbb{P}(Z > z) \quad & \text{if } \ H_A\colon \mu > \mu_0
\end{aligned}
$$

4. Make a decision according to the significance level α chosen.
    - If $p < \alpha$, we *reject* the null hypothesis (and accept the alternative)
    - If $p > \alpha$, we *do not reject* the null hypothesis

Example: Student Rent (1)

A property developer claims that the average yearly rent per room in student accommodation is €5,000, with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5,187. Do the sample results support the developer's claim? (use $\alpha = 0.05$)

Example: Student Rent (1)

A property developer claims that the average yearly rent per room in student accommodation is €5,000, with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5,187. Do the sample results support the developer's claim? (use $\alpha = 0.05$)

1. $H_0 \colon \mu = 5000 \quad \text{vs} \quad H_A \colon \mu \neq 5000$

Example: Student Rent (1)

A property developer claims that the average yearly rent per room in student accommodation is €5,000, with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5,187. Do the sample results support the developer's claim? (use $\alpha = 0.05$)

1. $H_0: \mu = 5000 \quad \text{vs} \quad H_A: \mu \neq 5000$
2. Compute z:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{5{,}187 - 5{,}000}{735/\sqrt{50}} = 1.799.$$

Example: Student Rent (1)

A property developer claims that the average yearly rent per room in student accommodation is €5,000, with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5,187. Do the sample results support the developer's claim? (use $\alpha = 0.05$)

1. $H_0 \colon \mu = 5000 \quad \text{vs} \quad H_A \colon \mu \neq 5000$
2. Compute z:
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{5{,}187 - 5{,}000}{735/\sqrt{50}} = 1.799.$$
3. Compute p-value:
$$p = \mathbb{P}(Z < -1.799) + \mathbb{P}(Z > 1.799) = 2 \times 0.036 = 0.072\,.$$

Example: Student Rent (1)

A property developer claims that the average yearly rent per room in student accommodation is €5,000, with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5,187. Do the sample results support the developer's claim? (use $\alpha = 0.05$)

1. $H_0 : \mu = 5000 \quad \text{vs} \quad H_A : \mu \neq 5000$
2. Compute z:
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{5{,}187 - 5{,}000}{735/\sqrt{50}} = 1.799.$$
3. Compute p-value:
$$p = \mathbb{P}(Z < -1.799) + \mathbb{P}(Z > 1.799) = 2 \times 0.036 = 0.072\,.$$
4. As $p > 0.05$, we cannot reject H_0 at the 5% significance level.

Example: Student Rent (2)

Now suppose that the property developer claims that the average yearly rent is at most €5,000 per year, with a std of €735. How would you change your null/alternative hypotheses to test this claim? Carry out the test at significance $\alpha = 0.05$, with the same data as before.

Example: Student Rent (2)

Now suppose that the property developer claims that the average yearly rent is at most €5,000 per year, with a std of €735. How would you change your null/alternative hypotheses to test this claim? Carry out the test at significance $\alpha = 0.05$, with the same data as before.

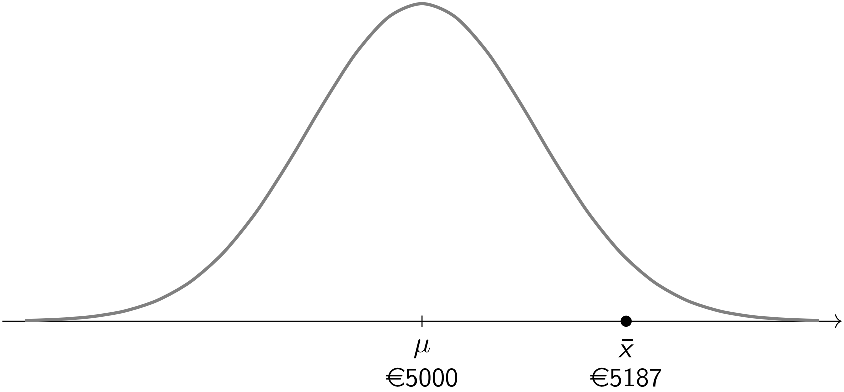
1. $H_0 \colon \mu = 5000 \quad \text{vs} \quad H_A \colon \mu > 5000$
2. z-score as before $z = 1.799$.
3. This time, since test is one-tail, we compute p-value as:

$$p = \mathbb{P}(Z > 1.799) = 0.036\,.$$

4. As $p < 0.05$, we reject H_0 and conclude that the average rent is higher than €5000 (at the 5% significance level).

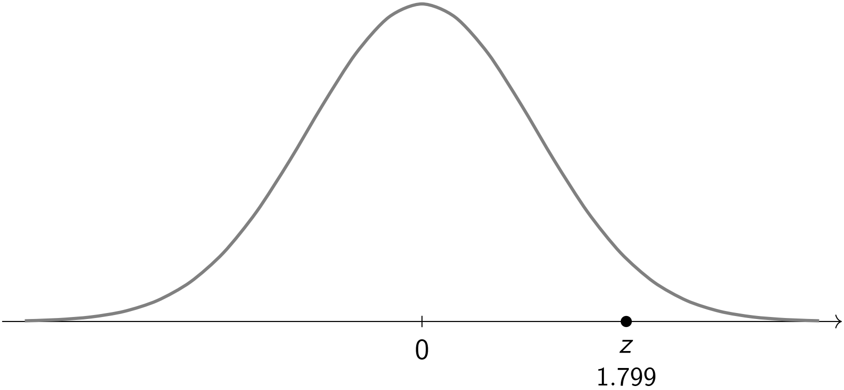
Student Rent: Visualisation

Start with original data: μ and $\bar{x}$

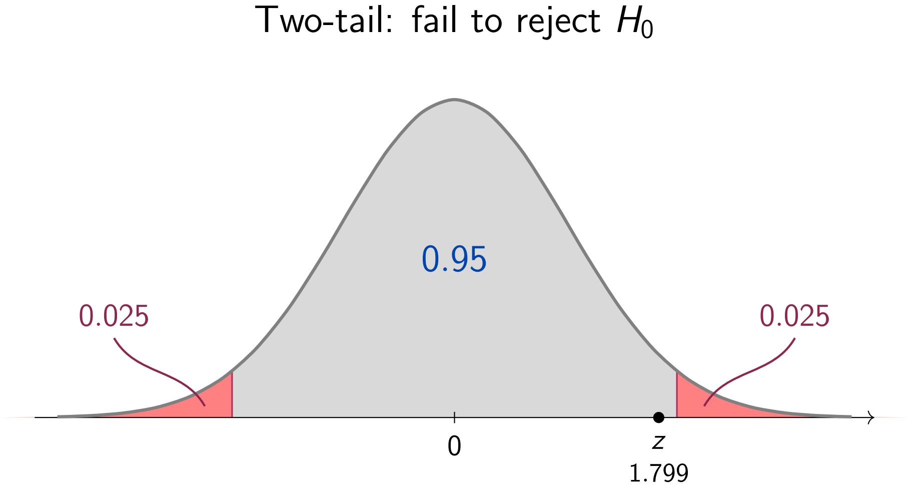


Student Rent: Visualisation

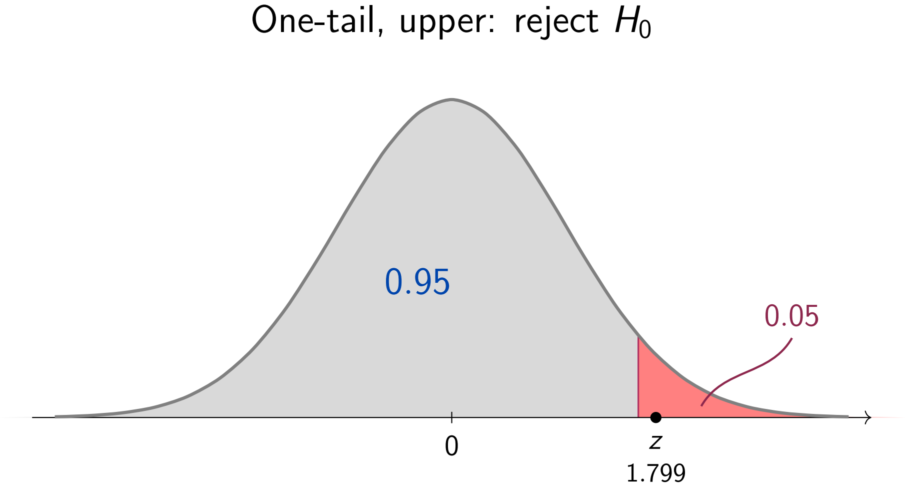
Standardise data $\rightarrow$ z-score



Student Rent: Visualisation



Student Rent: Visualisation



Rejection regions and p-values

- The **rejection region** $\mathcal{R}$ is the region of test-statistic values leading to a rejection of H_0. It depends on the alternative hypothesis.

$$H_A\colon \mu \neq \mu_0 \quad \to \quad \mathcal{R} = \left\{z < -z_{\alpha/2}\right\} \cup \left\{z > z_{\alpha/2}\right\}$$

$$H_A\colon \mu > \mu_0 \quad \to \quad \mathcal{R} = \{z > z_\alpha\}$$

$$H_A\colon \mu < \mu_0 \quad \to \quad \mathcal{R} = \{z < -z_\alpha\}$$

  where

$$\mathbb{P}(Z > z_\gamma) = \gamma, \quad Z \sim N(0,1).$$

- Checking whether the test-statistic falls in the rejection region is equivalent to checking whether the p-value is lower than the significance level α.

Example (for home): Road Usage

A road is being upgraded due to high heavy-freight traffic. The corporation reports an average of 71 trucks/hour (SD $= 12.2$), but engineers believe this is an underestimate. To test this, counts were taken over 50 one-hour periods, giving a sample mean of 74.1.

1. Test using a significance level of $\alpha = 0.1$.
2. Test using a significance level of $\alpha = 0.01$.

T-Tests for the Mean(s)

Student's t-test for the Population Mean

Recall the student rent example:

- hypothesised average rent of €5,000,
- sample average across 50 students of €5,187.

However, in reality, no-one tells you the std σ of the population that your sample comes from. How would you proceed?

Student's t-test for the Population Mean

Recall the student rent example:

- hypothesised average rent of €5,000,
- sample average across 50 students of €5,187.

However, in reality, no-one tells you the std σ of the population that your sample comes from. How would you proceed?

- In computing the test statistic (standardised mean)

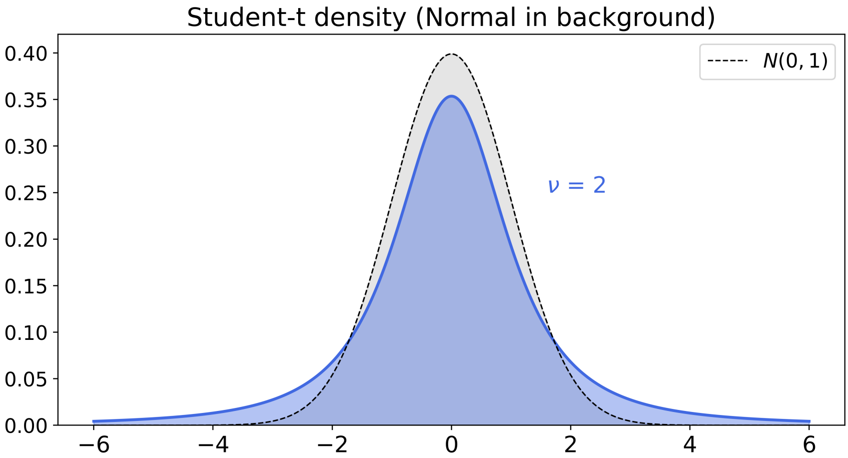
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

you can replace the unknown σ with the known *sample* standard deviation s.

- In this case, the sampling distribution of the test statistic is a Student-t distribution, under the null hypothesis $\mu = \mu_0$.

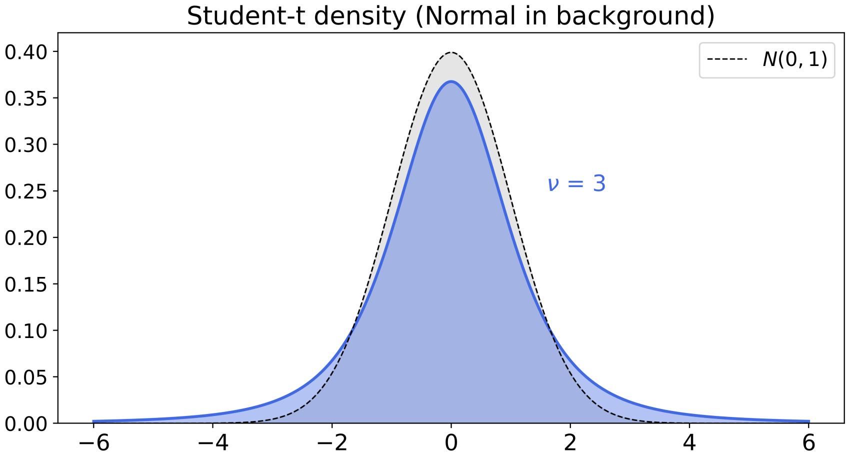
Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



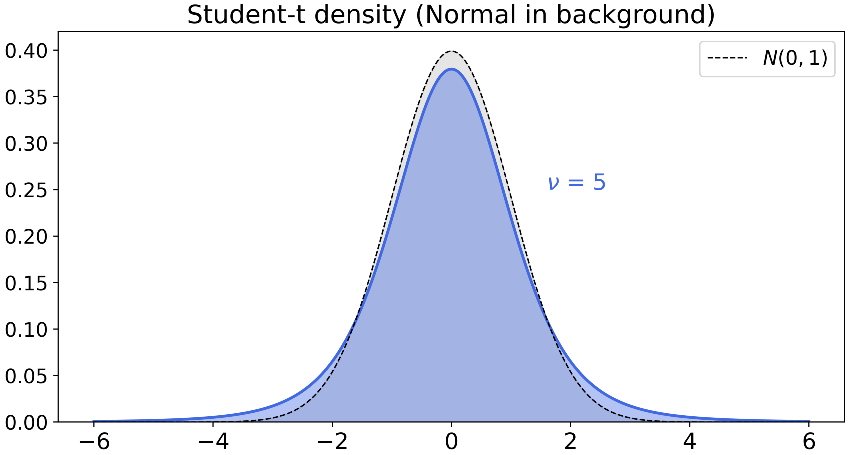
Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



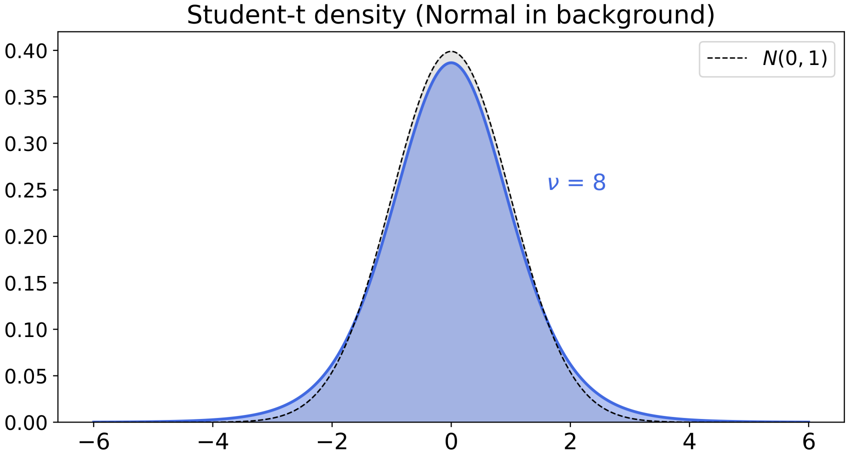
Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



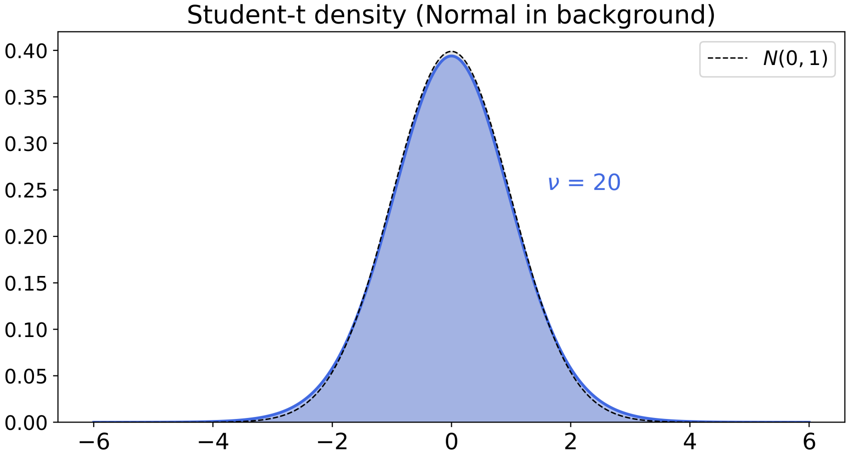
Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



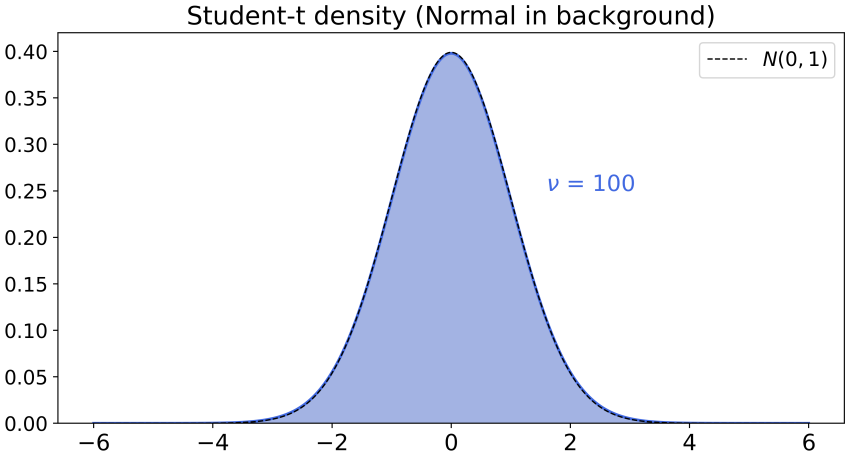
Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



Student-t Distribution

The t distribution depends on a parameter ν, called "degrees of freedom". It resembles a standard normal but it has *heavier tails than a normal*.



t-test for Population Mean μ

The test then works in a very similar way to the z-test.

t-test for Population Mean μ

The test then works in a very similar way to the z-test.

1. Specify the null and alternative hypotheses.

$$H_0: \mu = \mu_0 \qquad H_A: \begin{cases} \mu \neq \mu_0 & \text{two-tailed test} \\ \mu > \mu_0 & \text{one-tailed test (upper)} \\ \mu < \mu_0 & \text{one-tailed test (lower)} \end{cases}$$

t-test for Population Mean μ

The test then works in a very similar way to the z-test.

1. Specify the null and alternative hypotheses.

$$H_0 : \mu = \mu_0 \qquad H_A : \begin{cases} \mu \neq \mu_0 & \text{two-tailed test} \\ \mu > \mu_0 & \text{one-tailed test (upper)} \\ \mu < \mu_0 & \text{one-tailed test (lower)} \end{cases}$$

2. The test statistic in this situation is a t-score:

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

where the subscript $n-1$ indicates the degrees of freedom (df).
Under H_0, t_{n-1} follows a t-distribution with $n-1$ df.

t-test for Population Mean

3. Compute the p-value, according to the alternative hypothesis. Recall that the distribution of the test statistic is a Student-t distribution with $n-1$ df.

$$
\begin{aligned}
p &= \mathbb{P}(T_{n-1} < -|t_{n-1}|) + \mathbb{P}(T_{n-1} > |t_{n-1}|) && \text{if } \; H_A: \mu \neq \mu_0 \\
p &= \mathbb{P}(T_{n-1} > t_{n-1}) && \text{if } \; H_A: \mu > \mu_0 \\
p &= \mathbb{P}(T_{n-1} < t_{n-1}) && \text{if } \; H_A: \mu < \mu_0
\end{aligned}
$$

For this, you can use the t statistical tables.

4. As before, the threshold for the p-value depends on the significance level α.
   - If $p < \alpha$, we reject the null hypothesis and accept the alternative.
   - If $p > \alpha$, we fail to reject the null hypothesis.

Student's t-test: Key Characteristics Summary

- Uses the sample st.dev. s rather than the population st.dev. σ.
- One of the most commonly used statistical test.
- Test statistic follows a Student's t-distribution with $n-1$ df.
- Valid in particular for small samples.

  For large sample, t-distribution $\approx N(0,1)$ distribution, hence you can just use the standard normal tables to compute rejection regions and p-values.

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?
2. Compute the test statistic.
3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?
4. What about at a significance level of 0.01?

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?

$$H_0 \colon \mu = 100 \quad vs \quad H_A \colon \mu < 100.$$

2. Compute the test statistic.
3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?
4. What about at a significance level of 0.01?

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?
2. Compute the test statistic.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{97.5 - 100}{5/\sqrt{20}} = -2.236\,.$$

3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?
4. What about at a significance level of 0.01?

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?
2. Compute the test statistic.
3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?

   $p = \mathbb{P}(T_{19} < t) = 0.0188.$ Hence $p < 0.05 \Rightarrow$ reject H_0 at 5% significance

4. What about at a significance level of 0.01?

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?
2. Compute the test statistic.
3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?
4. What about at a significance level of 0.01?

$$p = 0.0188 > 0.01 \quad \Rightarrow \quad \text{cannot reject } H_0 \text{ at } 1\% \text{ significance}$$

Comparing Means

On some occasions (in fact, quite often!), we are interested in comparing two means with each other, rather than one mean to a reference value.

Examples

1. We want to test the efficiency of a medication in lowering blood pressure. A group of patients is administered the medication for two weeks. The blood pressure of each patient is measured before and after, and the two measurements are compared.
2. Are maths or stats students better at programming? We give the same programming task to two student groups (one from maths, one from stats) and compare the performance in the two groups.

Comparing Means

The two cases correspond to two different settings, which need (fairly) different approaches.

1. In the first case, we only have one group (the patients), whose "performance" is evaluated in two different conditions (before/after treatment).

   We use a **paired t-test** to compare the means in the two conditions.

2. In the second case, we have two groups (maths students and stats students) and want to compare the mean of the same variable (performance at programming test) in the two groups.

   In this case, we use an **independent samples t-test**.

Paired t-test for $\mu_1 - \mu_2$: Example

This case is *truly* nothing new. Let's see this on the blood pressure example.

Suppose there are $n = 6$ patients in the sample, whose "before treatment" and "after treatment" measurement are as follows:

Before (X_1)	After (X_2)
78	69
100	92
76	68
82	85
85	80
107	95

Paired t-test for $\mu_1 - \mu_2$: Example

This case is *truly* nothing new. Let's see this on the blood pressure example.

Suppose there are $n = 6$ patients in the sample, whose "before treatment" and "after treatment" measurement are as follows:

Before (X_1)	After (X_2)
78	69
100	92
76	68
82	85
85	80
107	95

Call μ_i the true mean of X_i, for $i = 1, 2$.

Want to show that $\mu_2 < \mu_1$, so:

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 > \mu_2$$

All that counts is the **difference** between the two measurements, for each patient. On average, has there been a decrease?

Paired t-test for $\mu_1 - \mu_2$: Example

So, we consider the new variable $D = X_1 - X_2$:

Bef (X_1)	Aft (X_2)	Diff
78	69	-9
100	92	-8
76	68	-8
82	85	$+3$
85	80	-5
107	95	-12

The new variable has mean $\mu = \mu_1 - \mu_2$. In terms of D, the two hypotheses read:

$$\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_A : \mu_1 > \mu_2 \end{array} \quad \leftrightarrow \quad \begin{array}{l} H_0 : \mu = 0 \\ H_A : \mu > 0 \end{array}$$

We can then carry out a classical t-test on the sample D of the differences, assuming (and trying to reject!) that its mean is 0. The test statistic is:

$$t_{n-1} = \frac{\bar{x}_D - 0}{s_D / \sqrt{n}}.$$

Independent Samples t-test

- In this case we have two different groups, *ie* two different populations. We are interested in comparing the means of two populations, μ_1 and μ_2.
- Think of the average programming test score for stats/maths students.
- For each of the two populations, we have a sample:
  - Sample 1: size n_1, mean $\bar{x}_1$, st.dev. s_1
  - Sample 2: size n_2, mean $\bar{x}_2$, st.dev. s_2.
- Question: Is the observed difference in sample means significant to conclude that the two original means μ_1 and μ_2 differ?

Independent Samples t-test

1. Specify the null and alternative hypotheses.

$$H_0 \colon \mu_1 = \mu_2 \qquad H_A \colon \begin{cases} \mu_1 \neq \mu_2 & \text{two-tailed test} \\ \mu_1 > \mu_2 & \text{one-tailed test (upper)} \\ \mu_1 < \mu_2 & \text{one-tailed test (lower)} \end{cases}$$

Independent Samples t-test

1. Specify the null and alternative hypotheses.

$$H_0 : \mu_1 = \mu_2 \qquad H_A : \begin{cases} \mu_1 \neq \mu_2 & \text{two-tailed test} \\ \mu_1 > \mu_2 & \text{one-tailed test (upper)} \\ \mu_1 < \mu_2 & \text{one-tailed test (lower)} \end{cases}$$

2. The test statistic is:

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} .$$

Note that $df = n_1 + n_2 - 2$ and the expression of the "pooled" standard error at the denominator.

Independent Samples t-test

3. Compute the p-value, according to the alternative hypothesis H_A.

$$
\begin{aligned}
p &= \mathbb{P}(T < -|t|) + \mathbb{P}(T > |t|) && \text{if } H_A\colon \mu_1 \neq \mu_2 \\
p &= \mathbb{P}(T > t) && \text{if } H_A\colon \mu_1 > \mu_2 \\
p &= \mathbb{P}(T < t) && \text{if } H_A\colon \mu_1 < \mu_2
\end{aligned}
$$

For brevity, we have denoted as T a t-Student random variable with $n_1 + n_2 - 2$ df. The previous t-score $t_{n_1+n_2-2}$ has been denoted as t.

4. For significance level α:
   - If $p < \alpha$, we reject the null hypothesis and accept the alternative.
   - If $p > \alpha$, we fail to reject the null hypothesis.

Example: Manual Dexterity and Sport

A study investigated whether playing sport is associated with greater manual dexterity. Two groups of children were tested:

- 17 non-sport players, with sample mean 31.68 and sample st.dev 4.56 .
- 15 Sport players, with sample mean 32.19 and sample std.dev 4.34 .

Using a significance level $\alpha = 0.05$, test whether those who participate in sport have a higher manual dexterity, on average, than those who don't.

Example: Manual Dexterity and Sport

A study investigated whether playing sport is associated with greater manual dexterity. Two groups of children were tested:

- 17 non-sport players, with sample mean 31.68 and sample st.dev 4.56 .
- 15 Sport players, with sample mean 32.19 and sample std.dev 4.34 .

Using a significance level $\alpha = 0.05$, test whether those who participate in sport have a higher manual dexterity, on average, than those who don't.

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_A : \mu_1 < \mu_2$$

Example: Manual Dexterity and Sport

Available Data:

$$n_1 = 17, \bar{x}_1 = 31.68, \qquad s_1 = 4.56$$
$$n_2 = 15, \bar{x}_2 = 32.19, \qquad s_2 = 4.34$$

Pooled Standard Error: $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.5744.$

t-score:

$$t_{n_1+n_2-2} = t_{30} = \frac{\bar{x}_1 - \bar{x}_2}{SE} = -0.324\,.$$

***p*-value & Conclusion**: $p = \mathbb{P}(T_{30} < -0.324) = 0.37.$

Since $p > 0.05$, we fail to reject H_0: There is not enough evidence in the data to conclude that $\mu_2 > \mu_1$.

Types of Tests for the Mean: Summary

Overall Remarks

- Use a z-**test** when the population standard deviation σ is known.
- Use a t-**test** when σ is unknown (same test statistic, just replace σ with s).

Note: For large samples, the t-distribution $\approx N(0, 1)$.

When σ is Unknown

- **One-sample t-test:** compare one sample mean to a hypothesised mean.
- **Paired t-test:** same individuals measured twice (reduces to a one-sample t-test on differences).
- **Independent-samples t-test:** two separate groups, compare their means.

Confidence Intervals for the Mean: General Case

- For large samples the CLT applies, so the confidence interval for μ is:

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}.$$

- For small samples (assuming the population is normal): since σ is unknown and replaced by s, use the t-**distribution** instead of the z-distribution to find the p-value or the critical t-value.

- Thus, for small samples, the **$100(1-\alpha)\%$ confidence interval for μ** is:

$$\bar{x} \pm t_{n-1}(\alpha/2) \frac{s}{\sqrt{n}},$$

where $\mathbb{P}\big(T_{n-1} > t_{n-1}(\alpha/2)\big) = \frac{\alpha}{2}$. For example, $t_{12}(0.025) = 2.179$.